ZooKeys 1236: 1–17 (2025)
DOI: 10.3897/zookeys.1236.138465

Research Article

The first complete mitochondrial genome of a *Lethrus* species (Coleoptera, Geotrupidae) with phylogenetic implications

Réka Zsófia Bubán[1,2], Renáta Bőkényné Tóth[1], Csongor Freytag[1], Gábor Sramkó[2,3], Zoltán Barta[4,5], Nikoletta Andrea Nagy[4,5]

1 *One Health Institute, University of Debrecen, Nagyerdei krt. 98, H-4032 Debrecen, Hungary*
2 *Department of Botany, University of Debrecen, Egyetem tér 1, H-4032 Debrecen, Hungary*
3 *HUN-REN-UD Conservation Biology Research Group, University of Debrecen, Egyetem tér 1, H-4032 Debrecen, Hungary*
4 *Department of Evolutionary Zoology and Human Biology, University of Debrecen, Egyetem tér 1, H-4032 Debrecen, Hungary*
5 *HUN-REN-UD Behavioural Ecology Research Group, Department of Evolutionary Zoology, University of Debrecen, Egyetem tér 1, H-4032 Debrecen, Hungary*

Corresponding author: Nikoletta Andrea Nagy (nagy.nikoletta@science.unideb.hu)

Abstract

The flightless beetle genus *Lethrus* Scopoli, 1777 (Geotrupidae, Scarabaeoidea) has a large distribution area throughout Eurasia and is characterized by many species, especially in Middle Asia. Despite this diversity and the potential importance as models for speciation, *Lethrus* species are underrepresented in molecular databases. To fill this gap, we report the complete mitochondrial genome of *Lethrus scoparius* obtained using third-generation sequencing technology. The circular mitogenome is 24,944 bp long and has a structure characteristic of coleopterans. It contains 37 genes, including 13 protein-coding genes, two ribosomal RNA genes, 22 transfer RNA genes, and an A+T-rich non-coding region, the control region between the 12S rRNA and tRNA-*Ile* (GAU). The phylogenetic analysis of the superfamily Scarabaeoidea placed *L. scoparius* in the monophyletic family Geotrupidae, which is related to the family Scarabaeidae. The assembled mitochondrial genome is a valuable new genomic resource in the genus *Lethrus* and contributes to a better understanding of the evolutionary history of the genus and the entire family Geotrupidae.

Key words: Circular mitogenome, phylogenetic relationship, phytophagous, Scarabaeoidea, third-generation sequencing

OPEN ACCESS

Academic editor: Andrey Frolov
Received: 3 October 2024
Accepted: 3 March 2025
Published: 24 April 2025

ZooBank: https://zoobank.org/3B9BEE98-798A-4BBE-BA6A-0E81037E42F8

Citation: Bubán RZ, Tóth RB, Freytag C, Sramkó G, Barta Z, Nagy NA (2025) The first complete mitochondrial genome of a *Lethrus* species (Coleoptera, Geotrupidae) with phylogenetic implications. ZooKeys 1236: 1–17. https://doi.org/10.3897/zookeys.1236.138465

Copyright: © Réka Zsófia Bubán et al. This is an open access article distributed under terms of the Creative Commons Attribution License (Attribution 4.0 International – CC BY 4.0).

Introduction

The family Geotrupidae is a relatively small, monophyletic group within the large and diverse superfamily Scarabaeoidea (Coleoptera) (Vuts et al. 2014). The species belonging to this group are divided into three subfamilies: Geotrupinae, Taurocerastinae, and Lethrinae (Cunha et al. 2011). The latter group consists of only one genus, *Lethrus* Scopoli, 1777, which comprises about 130 species and is subdivided into several subgenera (Shapovalov 2022). These beetles are flightless and have fused elytra. Most species exhibit sexual dimorphism, in which the males have a pair of tusk-like appendages on the ventral side of their mandibles that the females do not bear (Bagaturov and Nikolajev

2015). Another distinctive feature of these species compared to their geotrupid relatives is that they are phytophagous, feeding on fresh plant material of various species (Shapovalov 2022).

Lethrus is one of the most species-rich genera within the family Geotrupidae, with a wide distribution range in the Palearctic, especially in semi-arid habitats from Central and Southeastern Europe to China and Mongolia (Král et al. 2013; Bagaturov and Nikolajev 2015; Shapovalov 2022). Although some species occur sympatrically, most *Lethrus* species have small allopatric distributions (e.g. Král et al. 2013; Shapovalov 2022). One possible explanation for this patchy distribution pattern is the flightless nature of these species that might have limited their dispersal ability and thus triggered an increased speciation rate within the genus, similar to other examples (Ikeda et al. 2012; Salces-Castellano et al. 2021), especially in the regions of Middle Asia and the Balkans (Scholtz 2000; Bagaturov and Nikolajev 2015; Tóth et al. 2019). Despite the relatively high number of species in the genus and the unique life history traits, there are few molecular resources available for this taxon, both for phylogenetic and ecological genetic purposes.

In a study aimed at elucidating the phylogenetic relationships between the European *Lethrus* species based on targeted mitochondrial and nuclear loci, molecular data on Middle Asian species included as outgroup taxa were presented for the first time (Tóth et al. 2019). They were placed at the basal position of the genus. In this study, one species with a limited distribution area, *Lethrus scoparius* Fischer von Waldheim, 1822, was found to represent a clade subordinate to the European species. This species lives in the western Tian Shan, where individuals inhabit the plains and foothills (Nikolajev 2003). Based on the morphology of the adults, *L. scoparius* shows remarkable differences from the other species of the genus and was therefore described as the only member of the subgenus *Autolethrus* and proposed for further molecular studies to clarify its relationship with other *Lethrus* species (Bagaturov and Nikolajev 2015).

Mitochondrial genomes are important resources for phylogenetic, phylogeographic and population genetic studies at broad taxonomic scales (Cameron 2014; Kern et al. 2020; Zhao et al. 2021) because they are relatively small, simple in structure, have conserved gene arrangements, stable compositions and are maternally inherited (Kern et al. 2020; Jia et al. 2023). Although complete, circular mitochondrial sequences are available for many Scarabaeoidea species, they are scarce in the family Geotrupidae. The sole mitochondrial genome from the genus is that of *Lethrus apterus* Laxmann, 1770 (Nagy et al. 2021), which is fully annotated but incomplete. *De novo* assemblies of mitogenomes using short reads in many cases resulted in partial genomes, without an assembled A+T-rich non-coding region, also known as the control region (CR), and even without genes located near this highly repetitive sequence (e.g. Song and Zhang 2018). With the advances of long-read sequencing platforms, the incompleteness of such assemblies can be addressed.

In this study, we used Oxford Nanopore Technologies (ONT) long-read sequencing to determine the complete mitochondrial genome of *Lethrus scoparius*. It is the first complete mitogenome of the genus and one of the two available circularly assembled mitogenomes in the family Geotrupidae. We describe the structure of the genome, codon usage, nucleotide composition, and gene order.

Additionally, we compare these features with the other complete mitochondrial genomes within the superfamily and examine the phylogenetic relationships of the taxa. The mitochondrial genome presented provides valuable information for future work to uncover the internal relationships within the genus *Lethrus* and at higher taxonomic levels.

Methods

Sample collection and DNA extraction

An adult male of *Lethrus scoparius* was collected on 1 May 2023 in Tian Shan, Kazakhstan (42°23.96'N, 70°27.66'E) and housed at the Department of Evolutionary Zoology and Human Biology, University of Debrecen, Hungary. The species was identified by morphological characteristics (Bagaturov and Nikolajev 2015) and stored in 96% ethanol at 4 °C until DNA extraction. High-molecular-weight (HMW) DNA was isolated from the whole body using a conventional DNA extraction method. Briefly, a lysis buffer with a final concentration of 3 mM CaCl2, 2% sodium dodecyl sulfate (SDS), 40 mM dithiothreitol (DTT), 250 µg/ml proteinase K, 100 mM Tris buffer (pH = 8.0) and 100 mM NaCl was used (Gilbert et al. 2007). Cell lysis lasted 2 h at 55 °C, during which the sample was inverted several times. After incubation, the sample was centrifuged at 14,000 *g* for 1 min and the supernatant was transferred to a clean 2 ml tube to avoid contamination with chitin in subsequent steps. Then the sample was incubated with 15 µl (10 mg/ml) RNase A (Roche, Switzerland) for 10 minutes at 37 °C. In the next step, 0.5 volume of 7.5 M ammonium acetate was added, and the sample was incubated at +4 °C for 10 min. After incubation, 0.5 volume of a chloroform-isoamyl alcohol mixture (24:1) was added and the sample was incubated for 10 min at room temperature. The sample was centrifuged at 10,000 *g* for 3 min, then the supernatant was transferred to a clean 2 ml tube. After this step, 1 volume of room temperature isopropanol was added, and the sample was incubated at +4 °C for 15 min. After the sample was centrifuged at 10,000 *g* for 3 min, the supernatant was carefully removed from the pellet. The pellet was washed twice with 1 volume of room temperature 70% ethanol and centrifuged at 10,000 *g* for 3 min in between. The pellet was air-dried at room temperature and then dissolved in 65 µl of 10 mM Tris-HCl (pH = 8.0).

The nucleic acid concentration and purity of the sample were measured using a NABI UV/Vis Nano Spectrophotometer (MicroDigital Co., Ltd, Korea). The integrity of DNA was checked by TBE agarose gel (1%) electrophoresis.

Residual RNA contamination was removed by an additional 30-min digestion step at 37 °C with RNase Cocktail™ Enzyme Mix (0.1 U RNase A and 4 U Rnase T1) (Invitrogen, USA). Further purification steps were performed with 0.6 volume of Ampure XP (Beckman Coulter, USA) according to the manufacturer's instructions. To remove short DNA fragments, 0.64 volume of Long Fragment Buffer (Oxford Nanopore Technologies, UK) were also added. The decontaminated DNA was dissolved in 20 µl of nuclease-free water (Omega Bio-tek, Inc., USA). The concentration and purity of DNA were checked using a NanoDrop One (Thermo Fisher Scientific Inc., USA), a Qubit 4 Fluorometer (Thermo Fisher Scientific Inc., USA) with the 1x dsDNA High Sensitivity Kit (Thermo Fisher Scientific Inc., USA) and TBE agarose gel electrophoresis (1%).

Library construction and sequencing

The third-generation sequencing library (ONT) was prepared from 1 μg of genomic DNA using the Ligation Sequencing Kit V14 (SQK-LSK114) (Oxford Nanopore Technologies, UK) according to the manufacturer's recommendations. For DNA end repair, the NEBNext® Companion Module for Oxford Nanopore Technologies® Ligation Sequencing (NEB, E7180S) (New England Biolabs. Inc.) was used. 52 ng of the final library was loaded two times onto an R10.4 flow cell (FLO-MIN114). The sequencing run was performed using a MinION Mk1C device for 72 h.

Genome assembly and annotation

The sequencing yielded 9.26 giga base pairs (Gbp) of raw data. The Super high-accuracy model with Guppy 6.5.7 (Wick et al. 2019) was used for basecalling. Raw sequencing reads have been deposited to the Sequence Read Archive (SRA) under the accession number SRR28464392. The quality of the long reads was checked with the R script of MinIONQC 1.4.2 (Lanfear et al. 2019). Reads of the used DNA control strand (DNA CS) were removed with NanoLyse 1.2.0, then NanoFilt 2.8.0 (De Coster et al. 2018) was used to trim both ends of the reads (--headcrop 50, --tailcrop 50) and eliminate reads with low quality (mean quality cut-off 8 (-q 8)) or insufficient length (shorter than 500 bp (-l 500)). The cleaned and filtered reads were analyzed with NanoPlot 1.40.0 (De Coster et al. 2018). The quality filtering resulted in 2,751,077 clean reads with a total length of 6.79 Gbp and a read N50 of 4,124 bp.

Minimap 2.17 (Li 2018) was used to map the filtered reads to the mitochondrial genome of *Lethrus apterus* (Nagy et al. 2021), the closest relative with a fully annotated but incomplete mitochondrial genome. Reads covering more than 70% of the reference mitogenome were accepted as mitochondrial sequences. *De novo* assembly of the mitochondrial reads was performed using Flye 2.9 (Kolmogorov et al. 2019) with the predicted genome size set to 16 kbp and the assembly coverage set to 100 (-g 16k --asm-coverage 100). The assembled sequence was polished using Racon 1.4.10 (Vaser et al. 2017) and medaka 1.7.2 (https://github.com/nanoporetech/medaka) with the model r1041_e82_400bps_sup_g615. The annotation of the assembled mitochondrial genome was performed with MITOS2 (Donath et al. 2019) based on the invertebrate genetic code. The annotated mitochondrial genome was visualized using Proksee (Grant et al. 2023).

The AT-, and GC-skew of the sequences were calculated with the following formulas: AT-skew = (A − T) / (A + T); and GC-skew = (G − C) / (G + C) (Perna and Kocher 1995). Amino acid abundances and relative synonymous codon usage (RSCU) values of mitochondrial protein-coding genes (PCGs) were analyzed with Ezcodon (Lee 2018) implemented in EZmito Server (Cucini et al. 2021) using the invertebrate mitochondrial genetic code.

Phylogenetic analysis

The complete circular and annotated mitochondrial genomes of the superfamily Scarabaeoidea were downloaded from NCBI GenBank on 13 January 2025 (Suppl. material 1: table S1). Species represented the families Lucanidae (three

subfamilies), Geotrupidae, Passalidae, Scarabaeidae (six subfamilies), and Trogidae. The mitogenome of the closest relative, *Lethrus apterus*, was also included, as it contains all coding regions needed to reconstruct the phylogenetic tree. In addition, the mitochondrial genomes of *Apatetica glabra* (Staphylinoidea), *Diamesus osculans* (Staphylinoidea), and *Sphaeridium bipustulatum* (Hydrophiloidea) were used as outgroup samples. The nucleotide sequences of PCGs and ribosomal RNA (rRNA) coding genes were collected from the annotations. The sequences of PCGs were aligned with MACSE 2.07 (Ranwez et al. 2018) using the invertebrate mitochondrial code (-gc_def 5), and MUSCLE 3.8.1551 (Edgar 2004) was used to align the rRNA gene sequences. The aligned nucleotide sequences were merged using AMAS.py (Borowiec 2016) to obtain a 13,812 bp long alignment. Phylogenetic reconstruction was performed using IQ-TREE 2.0.3 (Nguyen et al. 2015). All genetic regions were defined as distinct partitions and ModelFinder Plus (Kalyaanamoorthy et al. 2017) was used to find the best fitting partition scheme for the dataset. Branch support was assessed using the SH-like approximate likelihood ratio test (SH-aLRT) (Guindon et al. 2010) and the ultrafast bootstrap (UFBoot) (Minh et al. 2013) with 1000 replicates.

Results

General characteristics of the mitogenome

The assembled mitochondrial genome of *Lethrus scoparius* was complete, circular, and 24,944 bp long. The mitogenome contained 13 protein-coding genes (PCG), two ribosomal RNA genes (rRNA), 22 transfer RNA genes (tRNA), and the control region (CR) (Fig. 1, Table 1). Most of the PCGs and tRNA genes as well as the CR were encoded on the heavy strand, whereas the remaining four PCGs (*nad1*, *nad4*, *nad4l* and *nad5*), eight tRNA genes (tRNA-*Cys* (GCA), tRNA-*Gln* (UUG), tRNA-*His* (GUG), tRNA-*Leu* (UAG), tRNA-*Phe* (GAA), tRNA-*Pro* (UGG), tRNA-*Tyr* (GUA), tRNA-*Val* (UAC)), and the two rRNA genes were located on the light strand (Table 1).

The overall base composition of the mitogenome was 41.3% A, 13.6% C, 7.9% G, and 37.2% T, and the A+T content was 78.48%. The AT-skew was slightly positive: 0.052, and the GC-skew was negative: −0.268 (Suppl. material 1: table S2). Descriptive statistics of the protein-coding genes, transfer RNA genes and ribosomal RNA genes, including the amino acid frequency and the codon usage, length distribution and base content are detailed in the Suppl. material 1.

Non-coding and overlapping regions

The largest non-coding region in the mitogenome of *L. scoparius* was the control region at 9,752 bp in length. The CR was located between the 12S rRNA and the tRNA-*Ile* (GAU) genes and had an A+T content of 80.2%. The base content of this region was 42.04% A, 13.33% C, 6.45% G, and 38.18% T. The AT-skew of the CR had a positive value of 0.048 and the GC-skew had a negative value of −0.348. The second largest non-coding region in the mitogenome had a length of 499 bp and was located between the nad2 and tRNA-*Trp* (UCA) genes. In addition, 12 other intergenic spacers were found that were between 1 bp and 40 bp in length (Table 1).

Table 1. Annotation of the mitochondrial genome of *Lethrus scoparius*. The Strand column shows the orientation of the genes, where the + marks the heavy strand and – denotes the light strand. The ovl/nc column indicates the overlapping or intergenic (non-coding) nucleotides, where the positive value means intergenic nucleotides and the negative value indicates overlapping nucleotides.

Gene	Region (bp)	Strand	Length (bp)	ovl/nc	Codons	
					Start	Stop
tRNA-*Ile* (GAU)	1–67	+	67	0		
tRNA-*Gln* (UUG)	68–136	–	69	–1		
tRNA-*Met* (CAU)	136–204	+	69	30		
nad2	235–1,218	+	984	499	ATT	TAA
tRNA-*Trp* (UCA)	1,718–1,784	+	67	–8		
tRNA-*Cys* (GCA)	1,777–1,839	–	63	1		
tRNA-*Tyr* (GUA)	1,841–1,908	–	68	–2		
cox1	1,907–3,445	+	1,539	–5	TCG	TAA
tRNA-*Leu* (UAA)	3,441–3,504	+	64	–1		
cox2	3,504–4,187	+	684	3	ATG	TAA
tRNA-*Lys* (CUU)	4,191–4,261	+	71	0		
tRNA-*Asp* (GUC)	4,262–4,328	+	67	0		
atp8	4,329–4,484	+	156	–7	ATC	TAA
atp6	4,478–5,149	+	672	1	ATG	TAA
cox3	5,151–5,938	+	788	–1	ATG	TA(A)
tRNA-*Gly* (UCC)	5,938–6,003	+	66	0		
nad3	6,004–6,357	+	354	–2	ATA	TAG
tRNA-*Ala* (UGC)	6,356–6,421	+	66	–1		
tRNA-*Arg* (UCG)	6,421–6,486	+	66	3		
tRNA-*Asn* (GUU)	6,490–6,554	+	65	0		
tRNA-*Ser* (UCU)	6,555–6,621	+	67	0		
tRNA-*Glu* (UUC)	6,622–6,688	+	67	–2		
tRNA-*Phe* (GAA)	6,687–6,752	–	66	–3		
nad5	6,750–8,463	–	1,714	9	ATT	T(AA)
tRNA-*His* (GUG)	8,473–8,538	–	66	–1		
nad4	8,538–9,874	–	1,337	–7	ATG	TA(A)
nad4l	9,868–10,158	–	291	2	ATG	TAA
tRNA-*Thr* (UGU)	10,161–10,224	+	64	0		
tRNA-*Pro* (UGG)	10,225–10,288	–	64	2		
nad6	10,291–10,794	+	504	–1	ATT	TAA
cytb	10,794–11,936	+	1,143	–2	ATG	TAG
tRNA-*Ser* (UGA)	11,935–12,000	+	66	17		
nad1	12,018–12,929	–	912	40	ATA	TAA
tRNA-*Leu* (UAG)	12,970–13,030	–	61	–23		
16S rRNA	13,008–14,320	–	1,313	18		
tRNA-*Val* (UAC)	14,339–14,408	–	70	0		
12S rRNA	14,409–15,192	–	784	0		
CR	15,193–24,944	+	9,752	—		

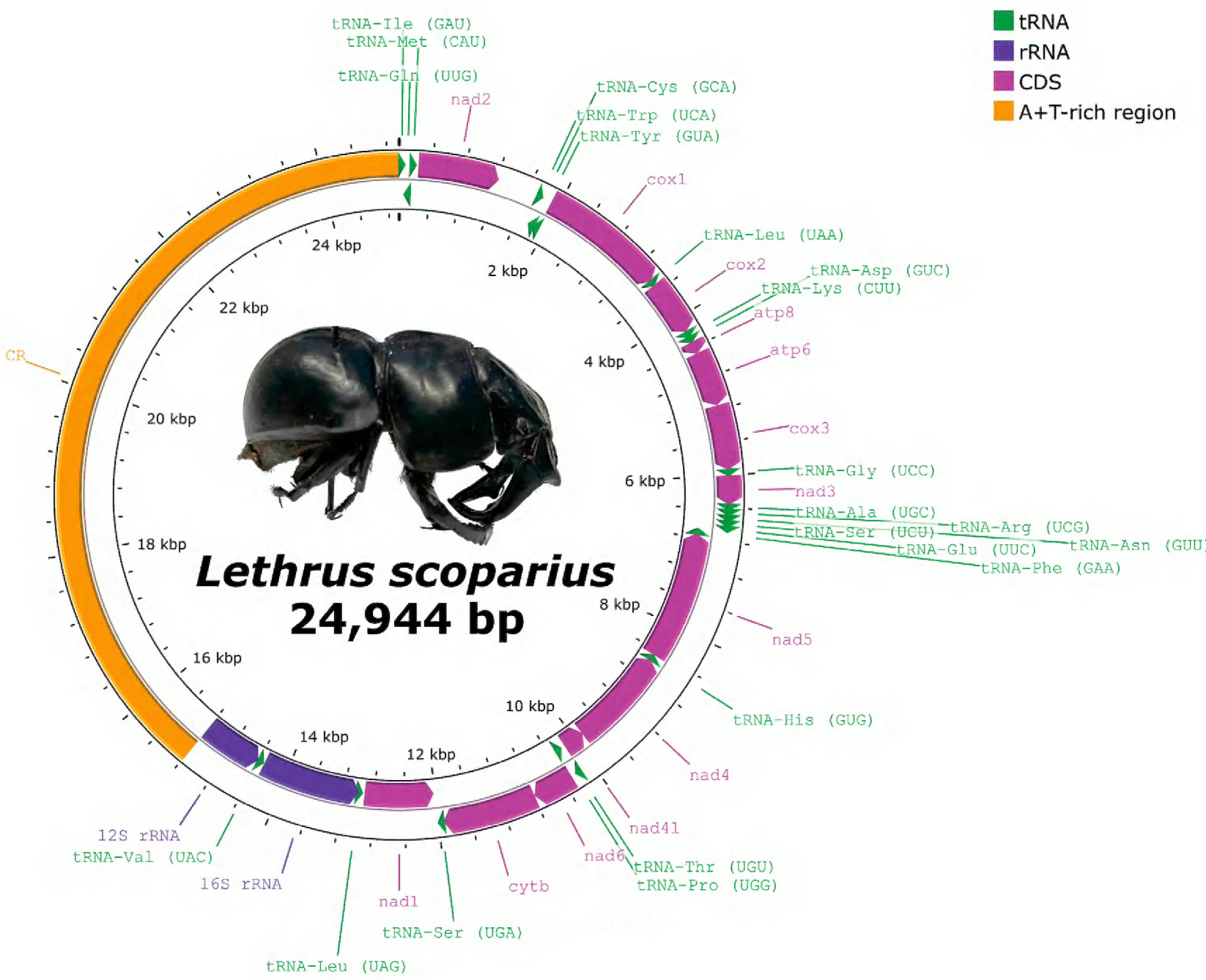

Figure 1. The circular mitochondrial genome map of *Lethrus scoparius*. Genes on the outer circle are located on the heavy strand, whereas the inner circle represents genes on the light strand. Protein-coding genes are shown in purple, tRNA genes in green, rRNA genes in blue, and the control region is represented in orange. Photo taken by RZB.

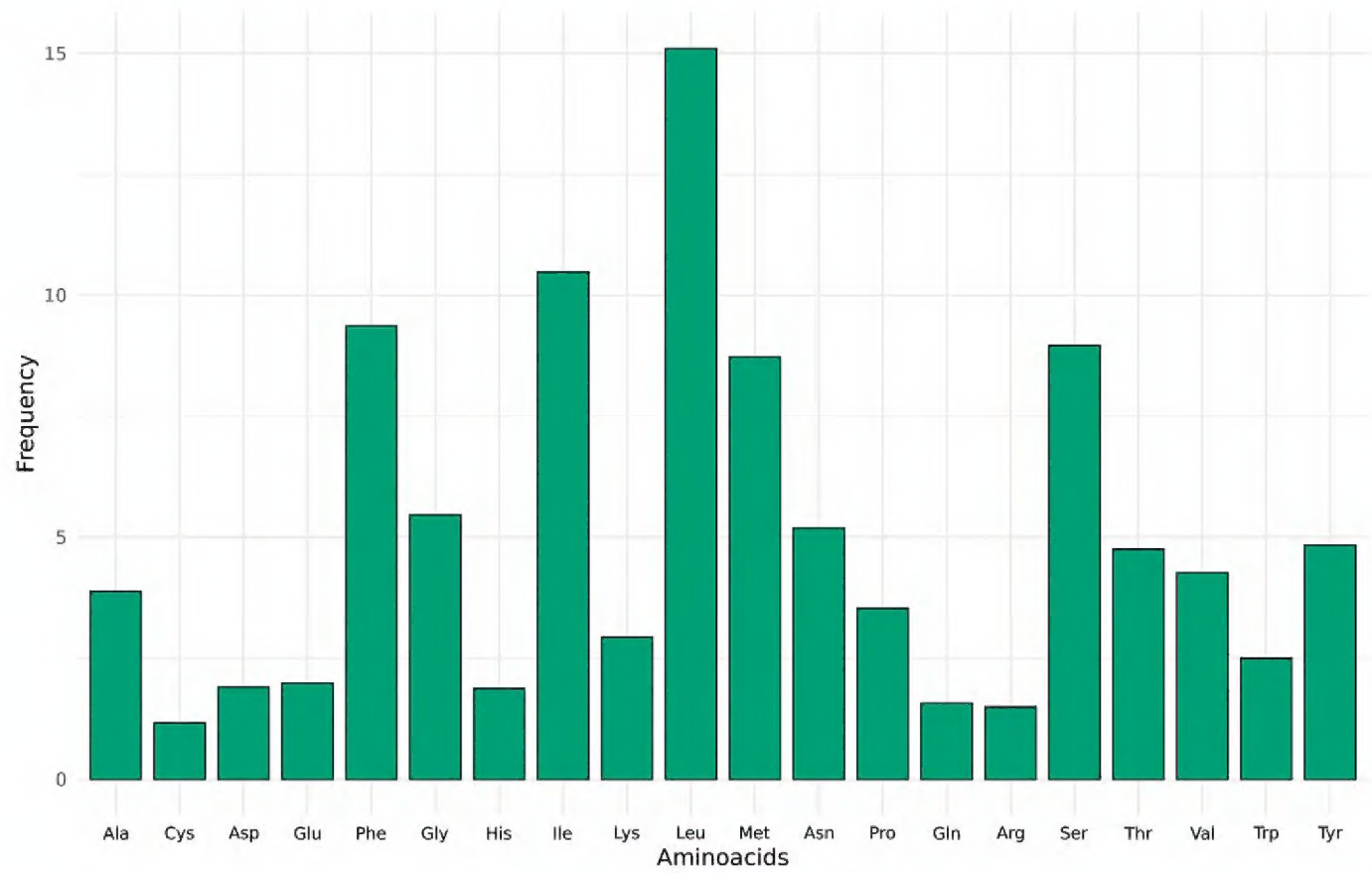

Figure 2. Amino acid frequency of the PCGs in the assembled mitochondrial genome of *Lethrus scoparius*. Three-letter amino acid code is shown on the X-axis and the frequency of amino acids in percentage on the Y-axis.

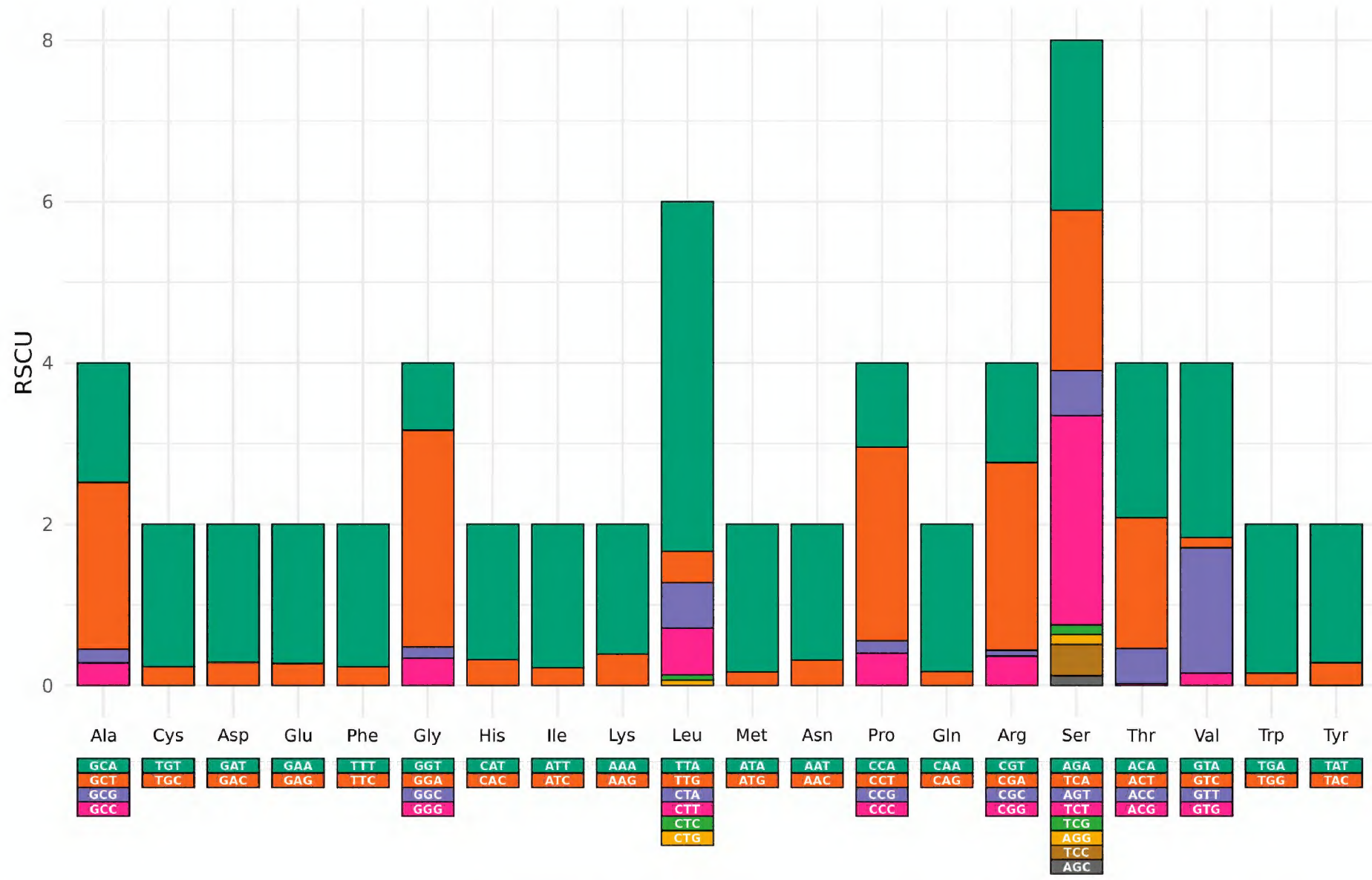

Figure 3. Relative synonymous codon usage (RSCU) of the PCGs in the assembled mitochondrial genome of *Lethrus scoparius*. Codon families and the three-letter amino acid codes are indicated on the X-axis and RSCU values on the Y-axis. Different colors of the bars represent different codons coding for the amino acid.

Several overlapping regions were observed in the mitogenome. The length of these sequences ranged from 1 bp to 23 bp (Table 1). The largest overlapping region was located between the tRNA-*Leu* (UAG) and the 16S rRNA genes. A 7-bp overlap with the "ATGATAA" motif was found between the PCGs *atp8* and *atp6*. In addition, an 8-bp overlap with the "AAGCCTTA" motif was detected between tRNA-*Trp* (UCA) and tRNA-*Cys* (GCA). Finally, an overlap with the "TTAACAT" motif was found between *nad4* and *nad4l*.

Phylogenetic analysis

We performed a maximum-likelihood (ML) phylogenetic tree reconstruction based on the nucleotide sequences of the PCGs and rRNA genes of 145 species of the superfamily Scarabaeoidea (Fig. 4). Our analysis revealed two monophyletic groups in addition to the outgroup, one containing the representatives of the family Lucanidae and the other consisting of the species of all four other families. Within the Lucanidae clade, all three subfamilies formed separate monophyletic clusters with the following branching: (Syndesinae (Aesalinae + Lucaninae)). The structure of the other clade was (Trogidae (Geotruipdae (Scarabaeidae + Passalidae))). Within the family Scarabaeidae, the subfamilies Rutelinae and Dynastinae were found to be monophyletic sister groups with the closest relationship to the Cetoniinae. The species of the Melolonthinae were clustered into a paraphyletic group with two main branches. One of these was a sister group to the clade

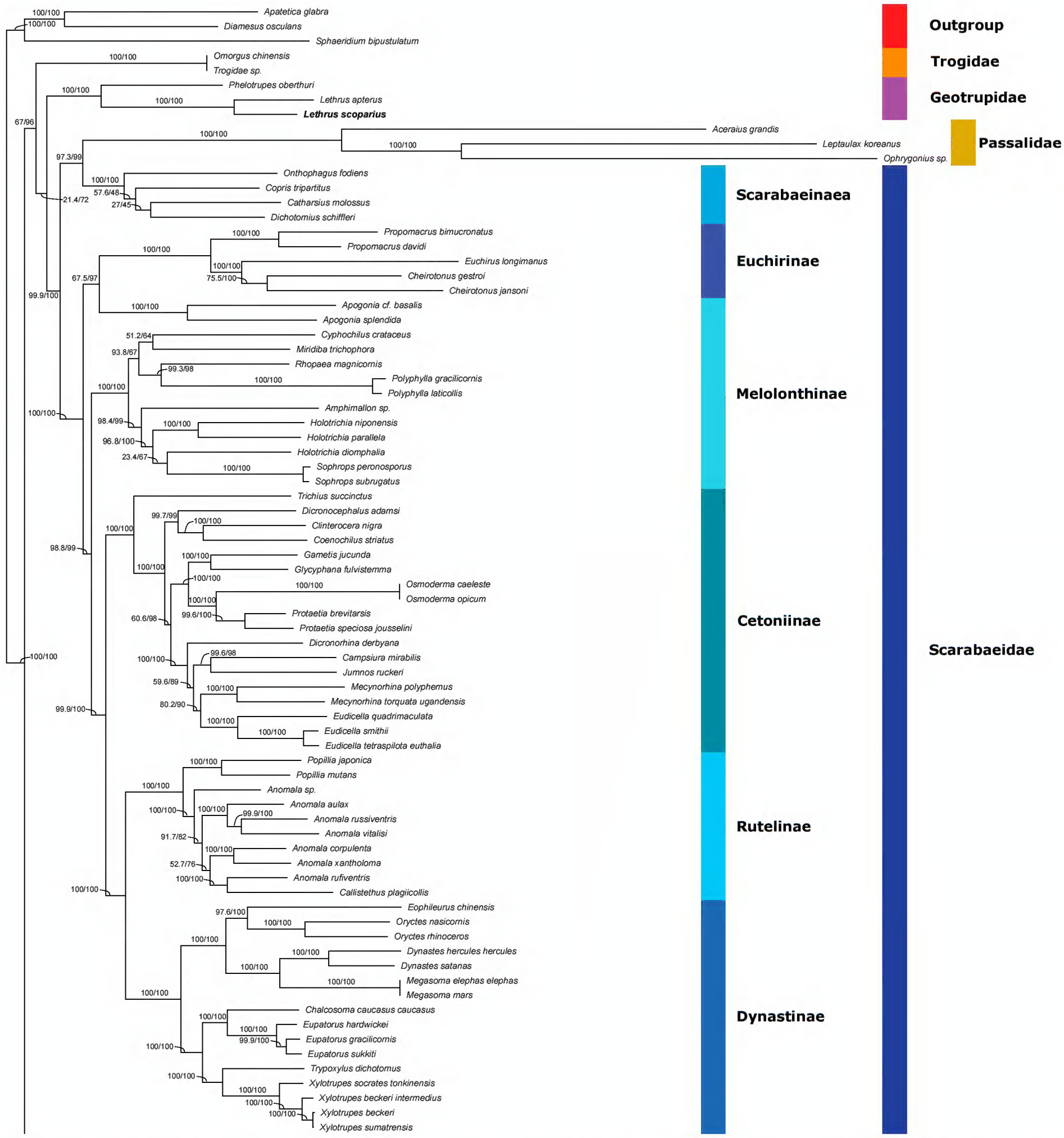

Figure 4. The reconstructed phylogenetic tree of the superfamily Scarabaeoidea based on species with a complete circular and annotated mitochondrial genome in the NCBI GenBank database (split into two parts). *Apatetica glabra*, *Diamesus osculans* and *Sphaeridium bipustulatum* were used as an outgroup. This maximum-likelihood tree was based on the nucleotide sequences of 13 mitochondrial PCGs and two rRNA genes. Branch supporting was evaluated with SH-like approximate likelihood ratio test (SH-aLRT) and ultrafast bootstrap (UFBoot) with 1000 replicates. Colors represent the respective subfamilies (left panel) and families (right panel). *Lethrus scoparius* is highlighted in bold.

((Rutelinae + Dynastinae) Cetoniinae), the other to the subfamily Euchirinae. Interestingly, the paraphyly of the family Scarabaeidae was a result of the cluster consisting of the subfamily Scarabaeinae and the family Passalidae. Within the family Geotrupidae, the two *Lethrus* species were found to have a sister relationship to the third representative of the family, *Phelotrupes oberthuri* Boucomont, 1905.

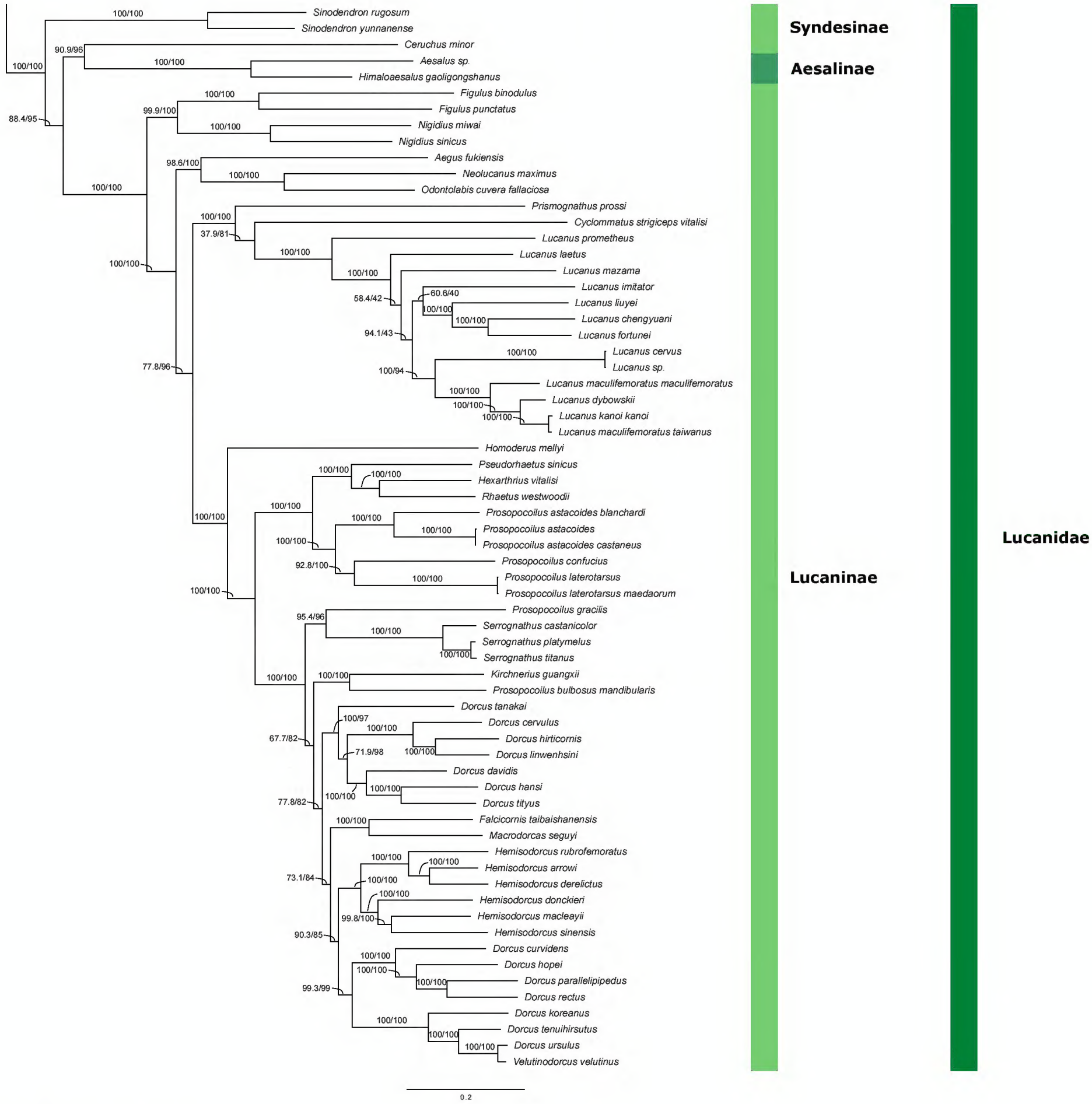

Figure 4. Continued.

Discussion

Lethrus scoparius is a rare Middle Asian earth-boring beetle from the genetically poorly characterized family Geotrupidae. Although *Lethrus* is one of the most species-rich genera in this family, molecular information on this scarabaeoid genus remains limited. In this study, we present the mitochondrial genome of *L. scoparius*, the first complete and well-characterized mitogenome of the genus. The mitogenome is 24,944 bp long and contains 37 genes (13 protein-coding genes, two ribosomal RNA genes and 22 transfer RNA genes) and a control region. The length of our assembled mitogenome is comparable with the elongated genomes found in Hercules beetles (*Dynastes*) (Morgan et al. 2022) and seed beetles (*Callosobruchus*) (Sayadi et al. 2017). The observed arrangement of genes was identical to the order of the insect ancestral mitochondrial genes (Sheffield

et al. 2008). The nucleotide sequence of the genome showed a characteristically high A+T content, as well as a positive AT-skew and a negative GC-skew.

We found several intergenic spacers and overlapping regions of different lengths. Of the observed overlaps, three motifs were proved to be conserved between different arthropod groups. The overlapping motif "ATGATAA" between PCGs *atp8* and *atp6* was also observed in other coleopteran mitogenomes (Wang and Tang 2017, 2018; Shi et al. 2023). In addition, an "AAGCCTTA" overlap motif was found between tRNA-*Trp* (UCA) and tRNA-*Cys* (GCA), similar to several other insect mitochondrial genomes (Wang and Tang 2017; Wang et al. 2018; Cao et al. 2019). Finally, the overlap of the "TTAACAT" motif between *nad4* and *nad4l* is characteristic of several other arthropod mitogenomes (Huang et al. 2015; Wang et al. 2016, 2018).

To reconstruct the phylogenetic relationships within the superfamily, we used the sequences of all Scarabaeoidea species with a publicly available complete, circular, and fully annotated mitochondrial genome at the date of completing this work. The dataset comprised 145 species resulting in currently the most comprehensive phylogenetic analysis of the superfamily. The basic structure of our reconstructed phylogenetic tree was similar to the phylogenetic relationships found by Ahrens et al. (2014) and Dietz et al. (2023). Previous studies investigating the relationships of the superfamily Scarabaeoidea mostly aimed to clarify the phylogeny within the Scarabaeidae and used members of the other families as outgroups (e.g. Gunter et al. 2016) or included only a few species from them (e.g. Ahrens et al. 2014; Song and Zhang 2018; Dietz et al. 2023). Within the family Scarabaeidae, the subfamily Melolonthinae appeared to be paraphyletic due to the presence of Euchirinae as sister to one of the Melolonthinae lineages. This is in line with the results of Ahrens et al. (2014) and Dietz et al. (2023), who found similar evolutionary relationships based on nuclear and mitochondrial markers. This supports the concept that the species forming the subfamily Euchirinae should be placed in the tribe Euchirini within Melolonthinae (Song and Zhang 2018; Ayivi et al. 2021; Guo et al. 2022). Although a recent study that published the mitogenome of the two *Propomacrus* species concluded that the subfamily Euchirinae forms a separate monophyletic cluster (Yi et al. 2024), the *Apogonia* species that caused the paraphyly within the subfamily Melolonthinae were not included in their analysis. Nevertheless, this taxonomic question would warrant further investigation with an improved set of markers to clarify the relationships between these two taxa.

Another interesting paraphyletic structure was found in our phylogenetic tree between the passalids and the Scarabaeidae. Although this exact relationship has not yet been hypothesized by others, the uncertain position of the family Passalidae is a known problem (see e.g. Dietz et al. 2023). All three members of this family that we included in our analysis were placed on extremely long branches, suggesting long-branch attraction bias – a common phenomenon in mitochondrial sequences (Timmermans et al. 2016). Other recent studies (Gunter et al. 2016; Guo et al. 2022; Dietz et al. 2023) have also placed representatives of the family on long branches, which could be addressed by increasing the number of passalid species with complete mitochondrial sequences.

Our results indicate that Geotrupidae and the clade of Scarabaeidae + Passalidae are sister groups. In the most recent studies, the family Geotrupidae has been considered to be the sister group to Scarabaeidae (Ahrens et al. 2014), Bolboceratidae (Song and Zhang 2018; Dietz et al. 2023), and Trogidae plus Glaphyridae

(Gunter et al. 2016). The results of Guo et al. (2022) show that the family Geotrupidae occupies different positions depending on the phylogenetic analysis method, although they only used this taxon as an outgroup in their study. Furthermore, the phylogenetic relationships within the family Geotrupidae have so far been analyzed mainly on the basis of larval morphology (Scholtz and Browne 1996; Verdú et al. 2004). Moreover, phylogenetic studies based on molecular markers have not yet included *Lethrus* species as representatives of the family Geotrupidae. Therefore, the mitochondrial sequence of *L. scoparius* is a useful contribution to corroborate the results found based on morphological characters.

The mitochondrial genome presented here is one of the few Scarabaeoidea mitogenomes generated using long-read sequencing technologies (Filipović et al. 2021; Morgan et al. 2022). The advantages of these methods could compensate for the shortcomings of the widely used short-read sequencing platforms. According to the results of Baeza et al. (2024), the latest Oxford Nanopore long-read sequencing chemical methods lead to a very low sequencing error rate, making the short-read polishing step in the assembly process redundant. With this method, the assembly of the non-coding control region, which often contains repetitive motifs, could be performed more accurately (Filipović et al. 2021). Elongated control regions reported in coleopteran species contain several tandem repeats, which have been proposed as informative markers for population genetic studies due to their higher evolutionary rate (Sayadi et al. 2017; Morgan et al. 2022). In addition, both Hercules beetles (Morgan et al. 2022) and seed beetles (Sayadi et al. 2017) have been shown to express specific sites of the control region using transcriptome sequencing data. Although the expression level was very low in both cases, these results suggest a role for these repeat sequences in a mitochondrial function. Our assembly is a novel example of the expansion of control regions in the Scarabaeoidea and suggests that it may be worthwhile to revisit the already assembled mitogenomes to detect expanded control regions and examine their tandem repeat content in the superfamily.

Conclusions

Among the members of the family Geotrupidae, our assembly of the mitogenome of *Lethrus scoparius* and the mitogenome of *Phelotrupes oberthuri* are the only representatives with circularly assembled and fully annotated complete mitochondrial genomes. This sequence provides valuable molecular resources for understanding the phylogeny of this family and the molecular evolution of the mitochondrial genome within the superfamily Scarabaeoidea, and a good example of the high efficiency of third-generation sequencing techniques in assembling mitochondrial genomes.

Acknowledgements

We sincerely thank M.F. Bagaturov for the taxonomic validation of the species investigated. We are thankful for the support of Juhász-Nagy Pál Doctoral School, University of Debrecen, Hungary. We thank for their assistance in fieldwork to Abidkulova Karime (Al-Farabi Kazakh National University, Almaty, Kazakhstan) and Jumanov Smatulla (Aksu-Zhabagly Nature Reserve, Zhabagly, Kazakhstan).

Additional information

Conflict of interest

The authors have declared that no competing interests exist.

Ethical statement

No ethical statement was reported.

Funding

N.A.N. was supported by the National Research, Development and Innovation Office (OTKA PD142602). Z.B. was supported by project no. TKP2021-NKTA-32 which has been implemented with the support provided by the Ministry of Culture and Innovation of Hungary from the National Research, Development and Innovation Fund, financed under the TKP2021-NKTA funding scheme. G.S. was supported by the Hungarian Ministry for Innovation and Technology via an NKFI-FK project (FK137962).

Author contributions

Réka Zsófia Bubán - Conceptualization, Data Curation, Formal analysis, Investigation, Writing – Original Draft, Writing – Review & Editing. Renáta Bőkényné Tóth – Methodology, Investigation, Resources, Writing – Review & Editing. Csongor Freytag – Methodology, Investigation, Resources, Writing – Review & Editing. Gábor Sramkó –Resources, Writing – Review & Editing. Zoltán Barta – Conceptualization, Resources, Writing – Review & Editing. Nikoletta Andrea Nagy – Conceptualization, Methodology, Data Curation, Formal analysis, Investigation, Funding acquisition, Writing – Original Draft, Writing – Review & Editing, Supervision.

Author ORCIDs

Csongor Freytag https://orcid.org/0000-0002-3356-4182
Gábor Sramkó https://orcid.org/0000-0001-8588-6362
Zoltán Barta https://orcid.org/0000-0002-7121-9865
Nikoletta Andrea Nagy https://orcid.org/0000-0001-7771-830X

Data availability

All of the data that support the findings of this study are available in the main text or Supplementary Information.

References

Ahrens D, Schwarzer J, Vogler AP (2014) The evolution of scarab beetles tracks the sequential rise of angiosperms and mammals. Proceedings of the Royal Society B: Biological Sciences 281(1791): 20141470. https://doi.org/10.1098/rspb.2014.1470

Ayivi SPG, Tong Y, Storey KB, Yu DN, Zhang JY (2021) The mitochondrial genomes of 18 new Pleurosticti (Coleoptera: Scarabaeidae) exhibit a novel trnQ-NCR-trnI-trnM gene rearrangement and clarify phylogenetic relationships of subfamilies within Scarabaeidae. Insects 12(11): 1025. https://doi.org/10.3390/insects12111025

Baeza JA, Minish JJ, Michael TP (2024) Assembly of mitochondrial genomes using nanopore long-read technology in three sea chubs (Teleostei: Kyphosidae). Molecular Ecology Resources 25(1): e14034. https://doi.org/10.1111/1755-0998.14034

Bagaturov MF, Nikolajev GV (2015) Overview of distribution of the genus *Lethrus* Scopoli, 1777 (Coleoptera: Geotrupidae). Caucasian Entomological Bulletin 11(2): 303–314. https://doi.org/10.23885/1814-3326-2015-11-2-303-314

Borowiec ML (2016) AMAS: a fast tool for alignment manipulation and computing of summary statistics. PeerJ 4: e1660. https://doi.org/10.7717/peerj.1660

Cameron SL (2014) Insect mitochondrial genomics: implications for evolution and phylogeny. Annual Review of Entomology 59: 95–117. https://doi.org/10.1146/annurev-ento-011613-162007

Cao JJ, Wang Y, Huang, YR, Li WH (2019) Mitochondrial genomes of the stoneflies *Mesonemoura metafiligera* and *Mesonemoura tritaenia* (Plecoptera, Nemouridae), with a phylogenetic analysis of Nemouroidea. ZooKeys 835: 43–63. https://doi.org/10.3897/zookeys.835.32470

Chen Q, Chen L, Liao CQ, Wang X, Wang M, Huang GH (2022) Comparative mitochondrial genome analysis and phylogenetic relationship among lepidopteran species. Gene 830: 146516. https://doi.org/10.1016/j.gene.2022.146516

Cucini C, Leo C, Iannotti N, Boschi S, Brunetti C, Pons J, Fanciulli PP, Frati F, Carapelli A, Nardi, F (2021) EZmito: a simple and fast tool for multiple mitogenome analyses. Mitochondrial DNA Part B 6(3): 1101–1109. https://doi.org/10.1080/23802359.2021.1899865

Cunha RL, Verdú JR, Lobo JM, Zardoya R (2011) Ancient origin of endemic Iberian earth-boring dung beetles (Geotrupidae). Molecular Phylogenetics and Evolution 59(3): 578–586. https://doi.org/10.1016/j.ympev.2011.03.028

De Coster W, D'Hert S, Schultz DT, Cruts M, Van Broeckhoven C (2018) NanoPack: visualizing and processing long-read sequencing data. Bioinformatics 34(15): 2666–2669. https://doi.org/10.1093/bioinformatics/bty149

Dietz L, Seidel M, Eberle J, Misof B, Pacheco TL, Podsiadlowski L, Ranasinghe S, Gunter NL, Niehuis O, Mayer C, Ahrens D (2023) A transcriptome-based phylogeny of Scarabaeoidea confirms the sister group relationship of dung beetles and phytophagous pleurostict scarabs (Coleoptera). Systematic Entomology 48(4): 672–686. https://doi.org/10.1111/syen.12602

Donath A, Jühling F, Al-Arab M, Bernhart SH, Reinhardt F, Stadler PF, Middendorf M, Bernt M (2019) Improved annotation of protein-coding genes boundaries in metazoan mitochondrial genomes. Nucleic Acids Research 47(20): 10543–10552. https://doi.org/10.1093/nar/gkz833

Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Research 32(5): 1792–1797. https://doi.org/10.1093/nar/gkh340

Filipović I, Hereward JP, Rašić G, Devine GJ, Furlong MJ, Etebari K (2021) The complete mitochondrial genome sequence of *Oryctes rhinoceros* (Coleoptera: Scarabaeidae) based on long-read nanopore sequencing. PeerJ 9: e10552. https://doi.org/10.7717/peerj.10552

Gilbert MTP, Moore W, Melchior L, Worobey M (2007) DNA extraction from dry museum beetles without conferring external morphological damage. PloS ONE 2(3): e272. https://doi.org/10.1371/journal.pone.0000272

Grant JR, Enns E, Marinier E, Mandal A, Herman EK, Chen C, Graham M, Van Domselaar G, Stothard P (2023) Proksee: in-depth characterization and visualization of bacterial genomes. Nucleic Acids Research 51(W1): W484–W492. https://doi.org/10.1093/nar/gkad326

Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O (2010) New algorithms and methods to estimate maximum-likelihood phylogenies: assessing

the performance of PhyML 3.0. Systematic Biology 59(3): 307–321. https://doi.org/10.1093/sysbio/syq010

Gunter NL, Weir T A, Slipinksi A, Bocak L, Cameron SL (2016) If dung beetles (Scarabaeidae: Scarabaeinae) arose in association with dinosaurs, did they also suffer a mass co-extinction at the K-Pg boundary?. PLoS ONE 11(5): e0153570. https://doi.org/10.1371/journal.pone.0153570

Guo S, Lin X, Song N (2022) Mitochondrial phylogenomics reveals deep relationships of scarab beetles (Coleoptera, Scarabaeidae). PLoS ONE 17(12): e0278820. https://doi.org/10.1371/journal.pone.0278820

Huang M, Wang Y, Liu X, Li W, Kang Z, Wang K, Li X, Yang D (2015) The complete mitochondrial genome and its remarkable secondary structure for a stonefly *Acroneuria hainana* Wu (Insecta: Plecoptera, Perlidae). Gene 557(1): 52–60. https://doi.org/10.1016/j.gene.2014.12.009

Ikeda H, Nishikawa M, Sota T (2012) Loss of flight promotes beetle diversification. Nature Communications 3(1): 648. https://doi.org/10.1038/ncomms1659

Jia W, Wei J, Niu M, Zhang H, Zhao Q (2023) The complete mitochondrial genome of *Aeschrocoris tuberculatus* and *A. ceylonicus* (Hemiptera, Pentatomidae) and its phylogenetic implications. ZooKeys 1160: 145–167. https://doi.org/10.3897/zookeys.1160.100818

Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS (2017) ModelFinder: fast model selection for accurate phylogenetic estimates. Nature Methods 14: 587–589. https://doi.org/10.1038/nmeth.4285

Kern EMA, Kim T, Park JK (2020) The mitochondrial genome in nematode phylogenetics. Frontiers in Ecology and Evolution 8: 250. https://doi.org/10.3389/fevo.2020.00250

Kolmogorov M, Yuan J, Lin Y, Pevzner PA (2019) Assembly of long, error-prone reads using repeat graphs. Nature Biotechnology 37: 540–546. https://doi.org/10.1038/s41587-019-0072-8

Král D, Hillert O, Drožová D, Šípek P (2013) *Lethrus* (*Lethrus*) *schneideri* sp. n. (Coleoptera, Geotrupidae) from Greece. ZooKeys 339: 93–106. https://doi.org/10.3897/zookeys.339.6132

Lanfear R, Schalamun M, Kainer D, Wang W, Schwessinger B (2019) MinIONQC: fast and simple quality control for MinION sequencing data. Bioinformatics 35(3): 523–525. https://doi.org/10.1093/bioinformatics/bty654

Lee BD (2018) Python implementation of codon adaptation index. Journal of Open Source Software 3(30): 905. https://doi.org/10.21105/joss.00905

Li H (2018) Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics 34(18): 3094–3100. https://doi.org/10.1093/bioinformatics/bty191

Minh BQ, Nguyen MAT, von Haeseler A (2013) Ultrafast approximation for phylogenetic bootstrap. Molecular Biology and Evolution 30(5): 1188–1195. https://doi.org/10.1093/molbev/mst024

Morgan B, Wang TY, Chen YZ, Moctezuma V, Burgos O, Le MH, Huang JP (2022) Long-read sequencing data reveals dynamic evolution of mitochondrial genome size and the phylogenetic utility of mitochondrial DNA in Hercules beetles (Dynastes; Scarabaeidae). Genome Biology and Evolution 14(10): evac147. https://doi.org/10.1093/gbe/evac147

Nagy NA, Rácz R, Rimington O, Póliska S, Orozco-terWengel P, Bruford MW, Barta Z (2021) Draft genome of a biparental beetle species, *Lethrus apterus*. BMC Genomics 22(1): 301. https://doi.org/10.1186/s12864-021-07627-w

Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ (2015) IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Molecular Biology and Evolution 32(1): 268–274. https://doi.org/10.1093/molbev/msu300

Nikolajev GV (2003) Zhuki-kravchiki (Scarabaeidae, Geotrupinae, Lethrini): Biologiya, Sistematika, Rasprostranenie, Opredelitel'. [*Lethrus*-beetles (Scarabaeidae, Geotrupinae, Lethrini): Biology, Systematics, Distribution, Keys.] Kazakh University, Almaty. 254 pp.

Perna NT, Kocher TD (1995) Patterns of nucleotide composition at fourfold degenerate sites of animal mitochondrial genomes. Journal of Molecular Evolution 41: 353–358. https://doi.org/10.1007/BF01215182

Ranwez V, Douzery EJP, Cambon C, Chantret N, Delsuc F (2018) MACSE v2: toolkit for the alignment of coding sequences accounting for frameshifts and stop codons. Molecular Biology and Evolution 35(10): 2582–2584. https://doi.org/10.1093/molbev/msy159

Salces-Castellano A, Andújar C, López H, Pérez-Delgado AJ, Arribas P, Emerson BC (2021) Flightlessness in insects enhances diversification and determines assemblage structure across whole communities. Proceedings of the Royal Society B 288(1945): 20202646. https://doi.org/10.1098/rspb.2020.2646

Sayadi A, Immonen E, Tellgren-Roth C, Arnqvist G (2017) The evolution of dark matter in the mitogenome of seed beetles. Genome Biology and Evolution 9(10): 2697–2706. https://doi.org/10.1093/gbe/evx205

Scholtz CH (2000) Evolution of flightlessness in Scarabaeoidea (Insecta, Coleoptera). Deutsche Entomologische Zeitschrift 47(1): 5–28. https://doi.org/10.1002/mmnd.4800470102

Scholtz CH, Browne DJ (1996) Polyphyly in the Geotrupidae (Coleoptera: Scarabaeoidea): a case for a new family. Journal of Natural History 30(4): 597–614. https://doi.org/10.1080/00222939600770311

Shapovalov AM (2022) New species of the genus *Lethrus* Scopoli, 1777 (Coleoptera: Geotrupidae: Lethrinae) from Fergana Valley, Kyrgyzstan. Zootaxa 5159(3): 414–424. https://doi.org/10.11646/zootaxa.5159.3.6

Sheffield NC, Song H, Cameron SL, Whiting MF (2008) A comparative analysis of mitochondrial genomes in Coleoptera (Arthropoda: Insecta) and genome descriptions of six new beetles. Molecular Biology and Evolution 25(11): 2499–2509. https://doi.org/10.1093/molbev/msn198

Shi F, Yu T, Xu Y, Zhang S, Niu Y, Ge S, Tao J, Zong S (2023) Comparative mitochondrial genomic analysis provides new insights into the evolution of the subfamily Lamiinae (Coleoptera: Cerambycidae). International Journal of Biological Macromolecules 225: 634–647. https://doi.org/10.1016/j.ijbiomac.2022.11.125

Song N, Zhang H (2018) The mitochondrial genomes of phytophagous scarab beetles and systematic implications. Journal of Insect Science 18(6): 11. https://doi.org/10.1093/jisesa/iey076

Timmermans MJ, Barton C, Haran J, Ahrens D, Culverwell CL, Ollikainen A, Dodsworth S, Foster PG, Bocak V, Vogler AP (2016) Family-level sampling of mitochondrial genomes in Coleoptera: compositional heterogeneity and phylogenetics. Genome Biology and Evolution 8(1): 161–175. https://doi.org/10.1093/gbe/evv241

Tóth JP, Bereczki J, Rácz R, Varga Z, Barta Z, Sramkó G (2019) Phylogenetic relationships in the genus *Lethrus* (Coleoptera: Geotrupidae) reveal contrasting evolutionary history in Europe. Systematic Entomology 44(4): 899–910. https://doi.org/10.1111/syen.12364

Vaser R, Sović I, Nagarajan N, Šikić M (2017) Fast and accurate de novo genome assembly from long uncorrected reads. Genome Research 27: 737–746. https://doi.org/10.1101/gr.214270.116

Verdú JR, Galante E, Lumaret JP, Cabrero-Sañudo FJ (2004) Phylogenetic analysis of Geotrupidae (Coleoptera, Scarabaeoidea) based on larvae. Systematic Entomology 29(4): 509–523. https://doi.org/10.1111/j.0307-6970.2004.00256.x

Vuts J, Imrei Z, Birkett MA, Pickett JA, Woodcock CM, Tóth M (2014) Semiochemistry of the Scarabaeoidea. Journal of Chemical Ecology 40: 190–210. https://doi.org/10.1007/s10886-014-0377-5

Wang Q, Tang G (2017) Genomic and phylogenetic analysis of the complete mitochondrial DNA sequence of walnut leaf pest *Paleosepharia posticata* (Coleoptera: Chrysomeloidea). Journal of Asia-Pacific Entomology 20(3): 840–853. https://doi.org/10.1016/j.aspen.2017.05.010

Wang Q, Tang G (2018) The mitochondrial genomes of two walnut pests, *Gastrolina depressa depressa* and *G. depressa thoracica* (Coleoptera: Chrysomelidae), and phylogenetic analyses. PeerJ 6: e4919. https://doi.org/10.7717/peerj.4919

Wang K, Li X, Ding S, Wang N, Mao M, Wang M, Yang D (2016) The complete mitochondrial genome of the *Atylotus miser* (Diptera: Tabanomorpha: Tabanidae), with mitochondrial genome phylogeny of lower Brachycera (Orthorrhapha). Gene 586(1): 184–196. https://doi.org/10.1016/j.gene.2016.04.013

Wang Y, Cao J, Murányi D, Li W (2018) Comparison of two complete mitochondrial genomes from Perlodidae (Plecoptera: Perloidea) and the family-level phylogenetic implications of Perloidea. Gene 675: 254–264. https://doi.org/10.1016/j.gene.2018.06.093

Wick RR, Judd LM, Holt KE (2019) Performance of neural network basecalling tools for Oxford Nanopore sequencing. Genome Biology 20: 129. https://doi.org/10.1186/s13059-019-1727-y

Yi C, Shu X, Wang L, Yin J, Wang Y, Wang Y, Zhang H, He Q, Zhao M (2024) The first report of complete mitogenomes of two endangered species of genus *Propomacrus* (Coleoptera: Scarabaeidae: Euchirinae) and phylogenetic implications. PLoS ONE 19(9): e0310559. https://doi.org/10.1371/journal.pone.0310559

Zhao W, Liu D, Jia Q, Wu X, Zhang H (2021) Characterization of the complete mitochondrial genome of *Myrmus lateralis* (Heteroptera, Rhopalidae) and its implication for phylogenetic analyses. ZooKeys 1070: 13–30. https://doi.org/10.3897/zookeys.1070.72742

Supplementary material 1

Supplementary data

Authors: Réka Zsófia Bubán, Renáta Bőkényné Tóth, Csongor Freytag, Gábor Sramkó, Zoltán Barta, Nikoletta Andrea Nagy

Data type: pdf

Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.

Link: https://doi.org/10.3897/zookeys.1236.138465.suppl1